

Step 1

Zorba xquery program to retrieve Guardian university data on architecture program. Uses REST call.

```

File Edit View Terminal Tabs Help
1 |:
2 file:      xget_g40.xq
3 date:      20-Apr-2009
4 author:     Gary Lewis
5 purpose:    Retrieve architecture info from the Guardian's university data.
6 reference:  http://www.guardian.co.uk/news/datablog/2009/mar/10/universityguide
7 OS command: zorba --indent -o xg40.xml -f -q xget_g40.xml
8 output:     xg40.xml
9 |:
10
11 import module namespace zorba-rest =
12   "http://www.zorba-xquery.com/zorba/rest-functions";
13 let $x := zorba-rest:get("http://spreadsheets.google.com/feeds/list/phNtm3LmDZEM6HUHUnVkPaA/odu/
  public/basic",
14   <payload>
15     <part name="gid">40</part>
16   </payload>)
17 return $x

```

1,1 All

Sample output of the XML returned. Edited to highlight important features.

```

File Edit View Terminal Tabs Help
1 <?xml version="1.0" encoding="UTF-8"?>
2 <zorba-rest:result xmlns:zorba-rest="http://www.zorba-xquery.com/zorba/rest-functions">
3   <zorba-rest:status-code>200</zorba-rest:status-code>
4   ...
5   <zorba-rest:payload>
6     <feed xmlns="http://www.w3.org/2005/Atom" xmlns:openSearch="http://a9.com/-/spec/opensearchchr
  ss/1.0/" xmlns:gsx="http://schemas.google.com/spreadsheets/2006/extended">
7     ...
8     <title type="text">40 Architecture</title>
9     ...
10    <entry>
11      ...
12      <title type="text">Rank</title>
13      <content type="text">_cokwr: Name of Institution, _cpzh4: Subject Description, _crell: P
  rofile Links, _chk2m: Guardian score/100, _ciyn3: % Satisfied with Teaching, _ckd7g: Expenditure
  per student (FTE), _clrrx: Student:staff ratio, _cyevm: Career prospects, _cztg3: Value added s
  core/10, _d180g: Average Entry Tariff, _d2mkx: % Satisfied with Assessment</content>
14    </entry>
15    <entry>
16      ...
17      <title type="text">1</title>
18      <content type="text">_cokwr: Cambridge, _chk2m: 100, _ckd7g: 10, _clrrx: 13, _cztg3: 6,
  _d180g: 533</content>
19    </entry>
20    ...
21  </feed>
22 </zorba-rest:payload>
23 </zorba-rest:result>

```

1,7 All

Step 2

Zorba xquery program to extract only the architecture data needed for analysis.

```
File Edit View Terminal Tabs Help
1 |:
2 file:      xget_arch.xq
3 date:      20-Apr-2009
4 author:    Gary Lewis
5 purpose:   Convert zorba's REST payload of Guardian university data on
6           architecture program [xg40.xml] to a format for joining.
7 reference: http://ouseful.wordpress.com/2009/03/13/joining-data-from-the-guardian-data-store
           -student-satisfaction-data/
8 OS command: zorba --indent -o xarch.xml -f -q xget_arch.xq
9 output:    xarch.xml
10           Includes only institution name and percent of students in
11           program who are satisfied with the teaching they receive.
12 |:
13 import module namespace zorba-rest = "http://www.zorba-xquery.com/zorba/rest-functions";
14 declare default element namespace "http://www.w3.org/2005/Atom";
15
16 <arch>{
17   for $e in fn:doc("xg40.xml")//feed/entry[title != "Rank"]
18     let $c := fn:data($e/content),
19         $n := fn:tokenize($c, ", _"),
20         $instName := xs:string(fn:tokenize($n[1], ": ")[2]),
21         $pctSatisfied := xs:double(fn:tokenize($n[3], ": ")[2])
22   where matches($c, "_ciyn3")
23   return
24     <inst>
25       <name>{$instName}</name>
26       <data>{$pctSatisfied}</data>
27     </inst>
28 }</arch>
```

1,1 All

Sample output.

```
File Edit View Terminal Tabs Help
1 <?xml version="1.0" encoding="UTF-8"?>
2 <arch xmlns="http://www.w3.org/2005/Atom">
3   <inst>
4     <name>UCL</name>
5     <data>98</data>
6   </inst>
7   <inst>
8     <name>Bath</name>
9     <data>93</data>
10  </inst>
11  <inst>
12    <name>Newcastle</name>
13    <data>87</data>
14  </inst>
```

1,1 Top

Step 3

Repeat steps 1 and 2 for planning program. Note: I was never able to get the REST call to return the planning data from The Guardian. It always returned architecture data. Either I did something wrong or there's some kind of throttle on data access. It's only an experiment, so not terribly important. Since the planning data set was so small, I manually constructed the XML that would have resulted from step 2.

Step 4

Zorba xquery program to join the architecture and the planning data.

```
File Edit View Terminal Tabs Help
1 |:
2 file:      xjoin_arch_plan.xq
3 date:      20-Apr-2009
4 author:    Gary Lewis
5 purpose:   Join architecture and planning data for analysis.
6 OS command: zorba --indent -o xarch_plan.xml -f -q xjoin_arch_plan.xq
7 output:    xarch_plan.xml
8            Includes institution name, percent of students in
9            architecture satisfied with teaching, percent of students in
10           planning satisfied with teaching.
11 Note:     Only institutions with both planning and architecture
12           programs are included.
13 |:
14
15 declare default element namespace "http://www.w3.org/2005/Atom";
16
17 <results>{
18 for $a in fn:doc("xarch.xml")/arch/inst
19 for $p in fn:doc("xplan.xml")/plan/inst[name = $a/name]
20 order by $a/name
21 return
22   <inst>
23     <name>{data($a/name)}</name>
24     <arch>{data($a/data)}</arch>
25     <plan>{data($p/data)}</plan>
26   </inst>
27 }</results>
```

1,1 All

Sample output of the joined data.

```
File Edit View Terminal Tabs Help
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results xmlns="http://www.w3.org/2005/Atom">
3   <inst>
4     <name>Anglia Ruskin</name>
5     <arch>70</arch>
6     <plan>75</plan>
7   </inst>
```

1,1 Top

Step 5

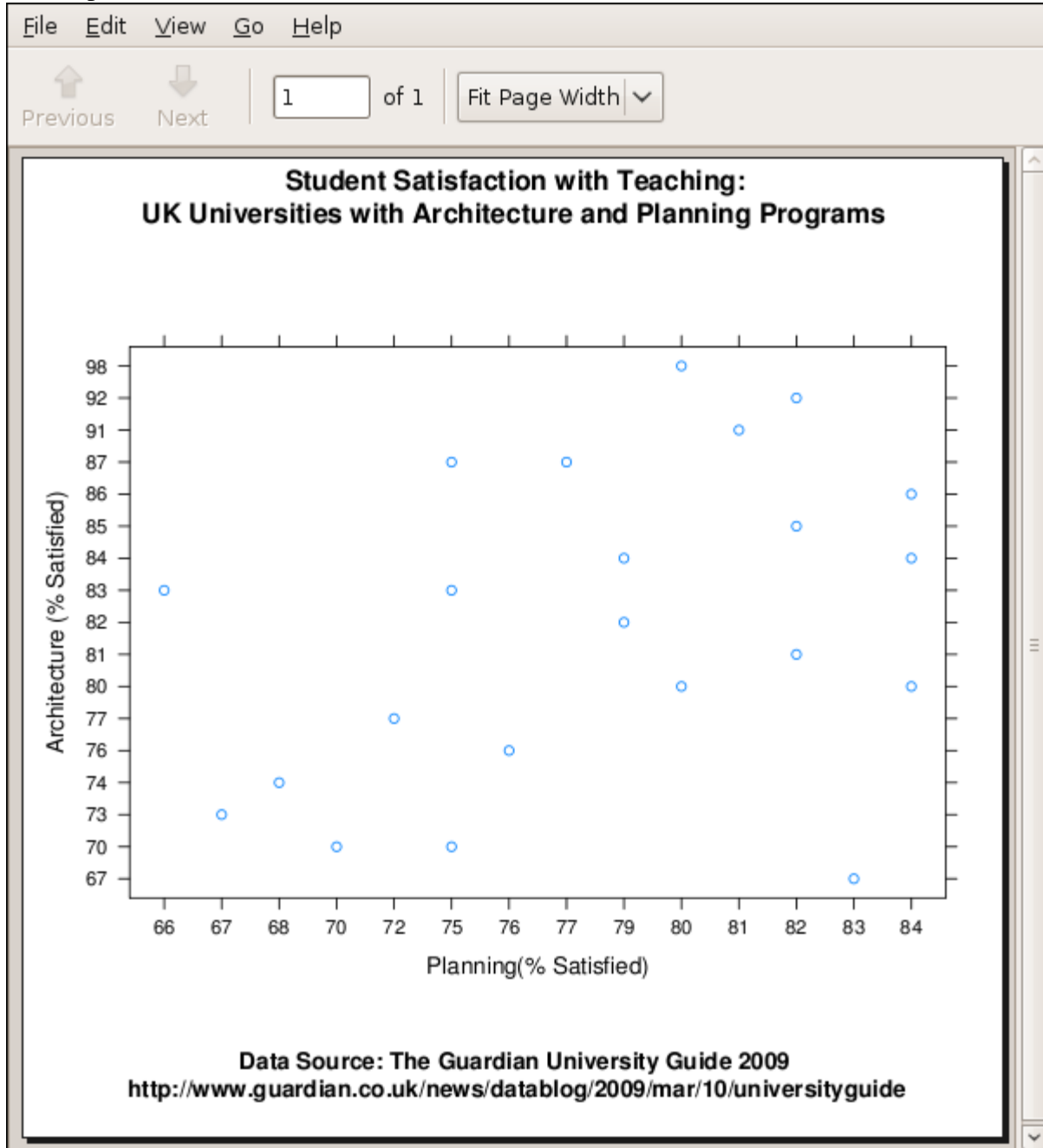
History of R statistics session to parse joined XML, create an R data frame, and produce a scatterplot.

```
File Edit View Terminal Tabs Help
1 ## file: xarch_plan.Rhistory
2 ## date: 20-Apr-2009
3 ## author: Gary Lewis
4 ## purpose: Parse XML data for architecture and planning. Produce a scatterplot, output to PDF.
5
6 library(XML)
7 library(lattice)
8 doc = xmlRoot(xmlTreeParse("xarch_plan.xml"))
9 tmp = xmlSApply(doc, function(x) xmlSApply(x, xmlValue))
10 tmp = t(tmp)
11 df = data.frame(tmp, row.names="name")
12 attach(df)
13 xyplot(arch~plan, data=df, ylab="Architecture (% Satisfied)", xlab="Planning(% Satisfied)\n\n",
main="Student Satisfaction with Teaching:\nUK Universities with Architecture and Planning Progra
ms\n\n", sub="Data Source: The Guardian University Guide 2009\n http://www.guardian.co.uk/news/d
atablog/2009/mar/10/universityguide\n")
:1 1,1 All
```

R data frame with architecture and planning data.

```
File Edit View Terminal Tabs Help
arch plan
Anglia Ruskin      70 75
Dundee             80 84
Glamorgan          76 76
Kingston           84 79
Leeds Met          74 68
Liverpool          81 82
Liverpool John Moores 73 67
London South Bank  80 80
Newcastle          87 75
Northumbria        85 82
Nottingham         84 84
Oxford Brookes    87 77
Plymouth           67 83
Portsmouth         86 84
Queen's, Belfast  77 72
Sheffield          91 81
Sheffield Hallam  92 82
Southampton Solent 70 70
UCL                98 80
Ulster             83 66
UWE Bristol        82 79
Westminster        83 75
>
```

Final scatterplot in R.



References

Tony Hirst, “Join'ing Data from the Guardian Data Store Student Satisfaction Data,” 13-March-2009. Available at: <http://ouseful.wordpress.com/2009/03/13/joining-data-from-the-guardian-data-store-student-satisfaction-data/>.

Donald MacLeod, “Get our full university data,” 10-March-2009. Available at: <http://www.guardian.co.uk/news/datablog/2009/mar/10/universityguide>.